Algorithmica (2001) 31: 442–457
DOI: 10.1007/s00453-001-0048-0

Algorithmica

Stochastic Analysis of Shell Sort

R. T. Smythe[1] and J. Wellner[2]

Abstract. We analyze the Shell Sort algorithm under the usual random permutation model. Using empirical distribution functions, we recover Louchard's result that the running time of the 1-stage of $(2, 1)$-Shell Sort has a limiting distribution given by the area under the absolute Brownian bridge. The analysis extends to $(h, 1)$-Shell Sort where we find a limiting distribution given by the sum of areas under correlated absolute Brownian bridges. A variation of $(h, 1)$-Shell Sort which is slightly more efficient is presented and its asymptotic behavior analyzed.

Key Words. Empirical distribution functions, Brownian bridge, Sorting algorithm, Random permutation model, Asymptotic distribution.

1. Introduction. Shell Sort is an algorithm that generalizes the method of sorting by insertion. It is essentially several stages of insertion sort. The algorithm was proposed in Shell (1959). The method received considerable attention over the past quarter of a century after Knuth's 1973 book popularized it (Knuth, 1973). From a practical point of view, the interest in Shell Sort stems from the fact that it is a rather practicable method of in situ sorting with little overhead and can be implemented with ease. From a theoretical standpoint, the interest is that insertion sort has an average of $\Theta(n^2)$ running time to sort n random keys, whereas the appropriate choice of the parameters of the stages of Shell Sort can bring down the order of magnitude. For instance, by a certain choice of the structure of the stages, a 2-stage Shell Sort can sort in $O(n^{5/3})$ average running time. Ultimately, an optimized choice of the parameter can come close to the information-theoretic lower bound of $O(n \ln n)$ average case.

The analysis of Shell Sort has stood as a formidable challenge. Most research on Shell Sort has gone in the direction of making good choices for the parameters of the stages to obtain good worst-case behavior (see, for example, the review paper of Sedgewick (1996)). We propose here to take the research along a different axis and to analyze the stochastic structure of the algorithm. We rederive the limit result of Louchard (1986) for $(2, 1)$-Shell Sort, which he proved by essentially combinatoric arguments, and we generalize the approach to the analysis of $(h, 1)$-Shell Sort. The integrated alsolute value of the Brownian bridge appears in the limiting distributions; the moments of the distribution of this random variable were found by Shepp (1982), and Johnson and Killeen (1983) gave an explicit characterization of the distribution function.

[1] Department of Statistics, Oregon State University, Corvallis, OR 97331, USA.
[2] Department of Statistics, University of Washington, Seattle, WA 98195, USA.

Received June 6, 2000; revised September 27, 2000. Communicated by K. Kemp and H. Prodinger.
Online publication August 9, 2001.

Section 2 gives a brief description of the algorithm; readers in search of more detail are advised to consult a source such as Sedgewick (1988), Sedgewick and Flajolet (1996), or Mahmoud (2000). In Section 3 the limiting distribution of the running time of (2, 1)-Shell Sort is derived using order statistics and empirical distribution functions. This is extended in Section 4 to $(h, 1)$-Shell Sort. Section 5 introduces a variation of $(h, 1)$-Shell Sort and gives the asymptotic distribution of its running time. In Section 6 we make some brief observations about $(h, k, 1)$-Shell Sort.

2. The Algorithm. To sort by insertion, one progressively adds keys to an already sorted file. This is achieved by identifying the rank of the next key by searching the available sorted list. The search can, of course, be done in many different ways. We restrict our attention to linear search, since it is a form that integrates easily into Shell Sort.

Shell Sort performs several stages of insertion sort and is well suited to arrays. We assume the data reside in a host linear array structure of size n. If the chosen Shell Sort uses k stages, a k-long integer sequence decreasing down to 1 is chosen to achieve a faster sort than plain insertion as follows. Suppose the sequence is $t_k, t_{k-1}, \dots, t_1 = 1$. In sorting n keys, the first stage sorts keys that are t_k positions apart in the list. Thus t_k subarrays of length at most $\lceil n/t_k \rceil$ each are sorted by plain insertion. In the second stage the algorithm uses the increment t_{k-1} to sort t_{k-1} subarrays of keys that are t_{k-1} positions apart (each subarray is of length at most $\lceil n/t_{k-1} \rceil$), and so on, down to the last stage, where an increment of 1 is used to insert-sort the whole array. Thus, in the last stage the algorithm executes plain insertion sort.

As an example, (2, 1)-Shell Sort sorts the array

$$3 \quad 2 \quad 6 \quad 5 \quad 9 \quad 8 \quad 1 \quad 4 \quad 7$$

in two stages. In the first stage the increments of 2 are used—the subarray of odd indexes is sorted by regular insertion sort, and the subarray of even indexes is sorted by regular insertion sort. The two interleaved arrangements

sorted odd positions:	1		3		6		7		9
sorted even positions:		2		4		5		8	

are then sorted by one run of the regular insertion sort on the whole array. The stage of Shell Sort that uses the increment t_j will be referred to as the t_j-stage of the algorithm. We use the usual notation $Z_{(j)}$ to denote the jth order statistic among $Z_1, \dots, Z_r$. (Technically, we should use $Z_{(j:\, r)}$ since $Z_{(j:\, r)}$ and $Z_{(j:\, s)}$ may differ for $s \neq r$; however, when the second subscript is understood, we drop it for simplicity.)

The notion of an *inversion* in a permutation is at the core of our analysis. Let $(\pi_1, \dots, \pi_n)$ be a permutation of $\{1, \dots, n\}$. One says that the pair (π_i, π_j) is an inversion if $\pi_i > \pi_j$, for $i < j$, that is when π_i and π_j are out of their natural order.

3. Analysis of (2, 1)-Shell Sort. Our goal in this paper is to discuss the stochastic behavior of Shell Sort when it operates on an array of n raw data. The usual probability model is the random permutation model, whereby the ranks of the data are equally likely to be any of the permutations of $\{1, \ldots, n\}$, each occurring with probability $1/n!$. This model is natural and is used as the standard for the analysis of sorting algorithms. The model covers a wide variety of real-world data models; for instance the entire class of data drawn from any continuous distribution follows the random permutation probability model. Henceforth the term *random* will specifically refer to data from this model. In this section we consider only $(2, 1)$-Shell Sort.

The difficulty in the stochastic analysis of Shell Sort lies in the fact that after the first stage the resulting data are no longer random. Instead, t_k sorted subarrays are interleaved. The second and subsequent stages may not then appeal to the results known for insertion sort. For example, the second stage of the $(2, 1)$-Shell's sort does not sort a random array of size n. The first stage somewhat orders the array as a whole, and many inversions are removed (some new ones may appear, though; see for example the positions of 5 and 6 before and after the 2-stage of our example).

Suppose our data are n real numbers from a continuous probability distribution. Because ordering is preserved by the probability integral transform, we may (and do) assume that the probability distribution is uniform on $(0, 1)$. We call the elements in odd positions X's and those in the even position Y's. Thus, if n is odd, initially our raw array prior to any sorting is

$$X_1, Y_1, X_2, Y_2, \ldots, Y_{\lfloor n/2 \rfloor}, X_{\lceil n/2 \rceil},$$

and if n is even the initial raw array is

$$X_1, Y_1, X_2, Y_2, \ldots, X_{n/2}, Y_{n/2}.$$

The 2-stage of the algorithm puts the X's in order among themselves and the Y's in order among themselves. Let S_n be the number of comparisons that $(2, 1)$-Shell Sort makes to sort n random keys, and let C_n be the number of comparisons that insertion sort makes to sort n random keys. The 2-stage of $(2, 1)$-Shell Sort makes two runs of insertion sort on the subarrays $X_1, \ldots, X_{\lceil n/2 \rceil}$ and $Y_1, \ldots, Y_{\lfloor n/2 \rfloor}$, thus requiring

$$C_{\lceil n/2 \rceil} \oplus \tilde{C}_{\lfloor n/2 \rfloor}$$

comparisons, where $\tilde{C}_j \stackrel{\mathcal{D}}{=} C_j$, and for all feasible i and j, C_i is independent of $\tilde{C}_j$.

The 1-stage now comes in, requiring additional comparisons to remove the remaining inversions. When we are about to insert $Y_{(j)}$, we place it among

$$\{X_{(1)}, \ldots, X_{(j)}\} \cup \{Y_{(1)}, \ldots, Y_{(j-1)}\}.$$

Because the 2-stage has sorted the Y's, $\{Y_{(1)}, \ldots, Y_{(j-1)}\}$ do not have any inversions with $Y_{(j)}$. Only $\{X_{(1)}, \ldots, X_{(j)}\}$ can introduce inversions. It is well known that the so-called sentinel version of insertion sort makes

$$C(\Pi_n) = n + I(\Pi_n)$$

comparisons to sort a permutation Π_n with $I(\Pi_n)$ inversions. Let V_j be the number of inversions $Y_{(j)}$ makes with all the elements that precede it, that is

$$V_j = \mathbf{1}_{\{X_{(1)} > Y_{(j)}\}} + \cdots + \mathbf{1}_{\{X_{(j)} > Y_{(j)}\}},$$

for $j = 1, \ldots, \lfloor n/2 \rfloor$. A symmetric argument applies to the insertion of $X_{(j)}$; define W_j as the number of inversions $X_{(j)}$ makes with all the elements that precede it:

$$W_j = \mathbf{1}_{\{Y_{(1)} > X_{(j)}\}} + \cdots + \mathbf{1}_{\{Y_{(j-1)} > X_{(j)}\}},$$

for $j = 1, \ldots, \lceil n/2 \rceil$. The number of inversions after the 2-stage is thus

$$I_n = \sum_{j=1}^{\lceil n/2 \rceil} V_j + \sum_{j=1}^{\lfloor n/2 \rfloor} W_j.$$

The 1-stage then requires $n + I_n$ comparisons

The overall number of comparisons S_n of (2, 1)-Shell Sort is therefore given by the convolution

$$S_n = C_{\lceil n/2 \rceil} \oplus \tilde{C}_{\lfloor n/2 \rfloor} \oplus n \oplus I_n. \tag{1}$$

Lent and Mahmoud (1996) developed Gaussian laws for the entire class of tree-growing search strategies, which includes linear search, the search method of Shell Sort. In particular, C_n, the number of comparisons that insertion sort performs to sort n random keys, is asymptotically normally distributed:

$$\frac{C_n - \frac{1}{4}n^2}{n^{3/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tfrac{1}{36}).$$

We focus on the limiting distribution of I_n. To avoid working with clumsy floors and ceilings, we assume n is even. The case n odd requires only minor adjustments that do not change the asymptotic result.

The next result is the key to our analysis:

LEMMA 1. *The total number of inversions after the* 2-*stage,* I_n, *has the representation*

$$I_n = \sum_{j=1}^{n/2} T_n^{(j)},$$

where

$$T_n^{(j)} = \left| \sum_{i=1}^{n/2} [\mathbf{1}_{\{Y_i < X_j\}} - \mathbf{1}_{\{X_i < X_j\}}] \right|.$$

PROOF. We have

$$\begin{aligned} I_n &= \sum_{j=1}^{n/2} (V_j + W_j) = \sum_{j=1}^{n/2} \left[\sum_{i=1}^{j-1} \mathbf{1}_{\{X_{(j)} < Y_{(i)}\}} + \mathbf{1}_{\{X_{(j)} > Y_{(j)}\}} + \cdots + \mathbf{1}_{\{X_{(j)} > Y_{(n/2)}\}} \right] \\ &= \sum_{j=1}^{n/2} \left| R_{(j)} - (2j - 1) \right|, \end{aligned}$$

where $R_{(j)}$ is the rank of $X_{(j)}$ among $X_1, \ldots, X_{n/2}, Y_1, \ldots, Y_{n/2}$. For $j = 1, 2, \ldots, n/2$, let

$$D_j \equiv \sum_{i=1}^{n/2} \mathbf{1}_{\{X_i \le X_j\}} = \sum_{i=1}^{n/2} \mathbf{1}_{\{X_i < X_j\}} + 1.$$

Then if R_j denotes the rank of X_j among all the X's and Y's, it follows that $R_{(D_j)} = R_j$ and

$$I_n = \sum_{j=1}^{n/2} \left| R_j - (2D_j - 1) \right|.$$

However,

$$R_j = \sum_{j=1}^{n/2} \mathbf{1}_{\{X_i < X_j\}} + \sum_{j=1}^{n/2} \mathbf{1}_{\{Y_i < X_j\}} + 1,$$

so

$$R_j - (2D_j - 1) \; = \; \sum_{j=1}^{n/2} [\mathbf{1}_{\{Y_i < X_j\}} - \mathbf{1}_{\{X_i < X_j\}}]. \qquad \square$$

The form in Lemma 1 can be expressed in terms of empirical distribution functions. Let $F_n(t)$ be the empirical distribution function of n i.i.d. random variables Z_i, that is

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{Z_i \le t\}}.$$

Then

$$T_n^{(j)} \overset{\text{a.s.}}{=} \frac{n}{2} \left| \hat{F}_{n/2}(X_j) - \left(\tilde{F}_{n/2}(X_j) - \frac{2}{n} \right) \right|,$$

where $\hat{F}_{n/2}$ and $\tilde{F}_{n/2}$ are the empirical distribution functions of the Y's and X's, respectively. Thus we may express I_n as

$$I_n = \frac{n}{2} \sum_{j=1}^{n/2} |\hat{F}_{n/2}(X_j) - \tilde{F}_{n/2}(X_j)| + o_p(n)$$

(where $o_p(n)$ denotes a term smaller than order n in probability, resulting from the "extra" $2/n$ in the expression for $T_n^{(j)}$), so that

$$\frac{I_n}{(n/2)^{3/2}} = \int_0^1 \sqrt{n/2} |\hat{F}_{n/2}(x) - \tilde{F}_{n/2}(x)| \, d\tilde{F}_{n/2}(x) + o_p(1). \tag{2}$$

The empirical process converges in the Skorohod topology on $D[0, 1]$, the space of right-continuous functions with left limits on $[0, 1]$, giving

$$\sqrt{n}\,(F_n(t) - t) \xrightarrow{\mathcal{D}} B(t),$$

where $B(t)$ is the (standard) Brownian bridge. Thus

$$\sqrt{n/2}|\hat{F}_{n/2}(x) - \tilde{F}_{n/2}(x)| \xrightarrow{\mathcal{D}} |B_1(x) - B_2(x)|, \tag{3}$$

where $B_1(t)$ and $B_2(t)$ are independent Brownian bridges. We note for future reference that the Brownian bridge is a Gaussian process, and $B_1 - B_2$ has the distribution of $\sqrt{2}*$ (a Brownian bridge), so

$$B_1(t) - B_2(t) \overset{\mathcal{D}}{=} \mathcal{N}(0, 2t(1-t)).$$

Also, the empirical measure $d\tilde{F}_{n/2}(x)$ converges a.s. to uniform measure on [0, 1], by the Glivenko–Cantelli theorem. Using the expression for I_n in (2), we have

$$\begin{aligned} &\int_0^1 \sqrt{n/2}|\hat{F}_{n/2}(x) - \tilde{F}_{n/2}(x)|\, d\tilde{F}_{n/2}(x) - \int_0^1 |B_1(x) - B_2(x)|\, dx \\ &\quad = \int_0^1 \sqrt{n/2}|\hat{F}_{n/2}(x) - \tilde{F}_{n/2}(x)|\, d\tilde{F}_{n/2}(x) - \int_0^1 |B_1(x) - B_2(x)|\, d\tilde{F}_{n/2}(x) \\ &\qquad + \int_0^1 |B_1(x) - B_2(x)|\, d\tilde{F}_{n/2}(x) - \int_0^1 |B_1(x) - B_2(x)|\, dx. \end{aligned} \tag{4}$$

For the first difference in (4), we can use the Skorokhod–Dudley–Wichura theorem (see, for example, Dudley (1997)) to construct, on a common probability space, versions $\hat{F}^*_{n/2}$, $\tilde{F}^*_{n/2}$, B^*_1, and B^*_2 of $\hat{F}_{n/2}$, $\tilde{F}_{n/2}$, B_1, and B_2 for which the convergence in (3) takes place uniformly a.s., so on this new space the first difference is bounded by

$$\sup_{\{0\le t\le 1\}} \sqrt{n/2}\left| |\hat{F}^*_{n/2}(t) - \tilde{F}^*_{n/2}(t)| - |B^*_1(t) - B^*_2(t)| \right|.$$

This converges to zero a.s. on the probability space of the construction. The Brownian bridge has a.s. bounded continuous paths, so the second difference also goes to zero a.s. by convergence of the empirical measure to uniform measure on [0, 1]. Thus on the original space, convergence in distribution is established, and we have the following result, first given by Louchard (1986):

THEOREM 1. *Let W_n be the number of comparisons made in the* 1*-stage of* (2,1)*-Shell Sort to sort n random keys, so that $W_n = n + I_n$. Then*

$$\frac{W_n}{(n/2)^{3/2}} \xrightarrow{\mathcal{D}} \sqrt{2}\int_0^1 |B(t)|\, dt,$$

where $B(t)$ is a standard Brownian bridge.

An infinite series for the c.d.f. of the limiting distribution of Theorem 1 is given by Johnson and Killeen (1983). The supremum of the uniform empirical process satisfies an exponential "tail" bound (Shorack and Wellner, 1986, p. 14) so it is easy to show from (2) that the convergence in distribution in Theorem 1 entails the convergence of moments of $W_n/(n/2)^{3/2}$ to those of $\sqrt{2}\int_0^1 |B(t)|\, dt$. These moments were derived by Shepp (1982); the limiting first and second moments for $(2, 1)$-Shell Sort had already been given by Knuth (1973).

COROLLARY 1.

(a) *The average number of comparisons made by the* 1*-stage of* (2,1)*-Shell Sort is asymptotically equivalent to* $(\sqrt{\pi}/8\sqrt{2})n^{3/2}$.
(b) *The variance of the number of comparisons is asymptotically equivalent to* $(7/240-\pi/128)n^3$.

Combining this result with the previously noted Gaussian limit for C_n, the number of comparisons made by the 2-stage of (2,1)-Shell Sort, we conclude that the asymptotic distribution of the total number of comparisons needed to sort n random keys is the convolution of a Gaussian law with the distribution of Theorem 1:

COROLLARY 2.

(a) *Let* S_n *be the number of comparisons made by* (2,1)*-Shell Sort to sort n random keys. Then*

$$\frac{S_n - \frac{1}{8}n^2}{n^{3/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tfrac{1}{144}) \oplus \frac{1}{2}\int_0^1 |B(t)|\, dt.$$

(b)

$$E(S_n) = \tfrac{1}{8}n^2 + \frac{\sqrt{\pi}}{8\sqrt{2}}n^{3/2} + o(n^{3/2}).$$

4. Analysis of (h, 1)-Shell Sort. For $h > 2$, we assume for simplicity that h divides n, the total number of keys; again, adjusting for this does not affect the limiting distribution. Prior to any sorting, our raw array is then given by

$$X_1^1, X_1^2, \ldots, X_1^h, X_2^1, \ldots, X_2^h, \ldots, X_{n/h}^1, \ldots, X_{n/h}^h$$

and the h-stage of the algorithm orders each of the h lists $\{X_j^1\}, \ldots, \{X_j^h\}$. The 1-stage then performs additional comparisons to order the now h-sorted list. When we are about to insert $X_{(j)}^r$, for $1 \le r \le h$, we place it among

$$\{X_{(1)}^1, \ldots, X_{(j)}^1\} \cup \cdots \cup \{X_{(1)}^{r-1}, \ldots, X_{(j)}^{r-1}\} \cup \{X_{(1)}^r, \ldots, X_{(j-1)}^r\} \\ \cup \cdots \cup \{X_{(1)}^h, \ldots, X_{(j-1)}^h\}.$$

As before, there will be no inversions produced with $\{X_{(i)}^r\}$, so the total number of inversions $X_{(j)}^r$ makes with the elements that precede it, for $1 \le j \le n/h$, is

$$\sum_{i=1}^{j}\sum_{k=1}^{r-1} \mathbf{1}_{\{X_{(i)}^k > X_{(j)}^r\}} + \sum_{i=1}^{j-1}\sum_{k=r+1}^{h} \mathbf{1}_{\{X_{(i)}^k > X_{(j)}^r\}}.$$

Let $r < s$ be any two integers between 1 and h. The total number of inversions encountered in the sort, which we denote I_n^h, is then the sum over all $\binom{h}{2}$ pairs (r, s) of

$$I_n^{r,s} \equiv \sum_{j=1}^{n/h} V_j^{r,s},$$

where

$$V_j^{r,s} \equiv \sum_{i=1}^{j-1} \mathbf{1}_{\{X^s_{(i)} > X^r_{(j)}\}} + \sum_{i=1}^{j} \mathbf{1}_{\{X^r_{(i)} > X^s_{(j)}\}}.$$

We denote the empirical process corresponding to the X_i^r by $F^r_{n/h}$, and let

$$U^r_{n/h}(t) \equiv \sqrt{n/h}(F^r_{n/h}(t) - t)$$

for $1 \le r \le h$. Arguing as in the previous section,

$$\frac{I_n^h}{(n/h)^{3/2}} = \sum_{r<s} \int_0^1 |U^r_{n/h}(t) - U^s_{n/h}(t)|\, dF^r_{n/h}(t) + o_p(1),$$

where the sum is over all pairs (r, s) with $1 \le r < s \le h$. The vector of processes $\underline{U}_{n/h} \equiv (U^1_{n/h}, \ldots, U^h_{n/h})$, by the classical Donsker theorem, satisfies

$$\underline{U}_{n/h} \xrightarrow{\mathcal{D}} \underline{B},$$

where $\underline{B} \equiv (B^1, \ldots, B^h)$ is a vector of independent Brownian bridge processes. Let

$$D^{r,s}_{n/h} \equiv U^r_{n/h}(t) - U^s_{n/h}(t) = F^r_{n/h}(t) - F^s_{n/h}(t).$$

It follows from the continuous mapping theorem that the $\binom{h}{2}$-vector of processes

$$\underline{D}_{n/h} \equiv (D^{1,2}_{n/h}, \ldots, D^{h-1,h}_{n/h})$$

satisfies

$$\underline{D}_{n/h} \xrightarrow{\mathcal{D}} \underline{D},$$

where $\underline{D} \equiv (D^{1,2}, \ldots, D^{h-1,h})$ and $D^{i,j} \equiv B^i - B^j$.

Next, we use the Skorokhod–Dudley–Wichura theorem as before to construct versions $\underline{\tilde{D}}_{n/h}$ and $\underline{\tilde{D}}$ of $\underline{D}_{n/h}$ and $\underline{D}$ on a common probability space such that $\underline{\tilde{D}}_{n/h}$ converges almost surely to $\underline{\tilde{D}}$ with respect to the supremum norm on $D[0, 1]^h$.

We can now establish the basic convergence result for $(h, 1)$-Shell Sort.

THEOREM 2. *Let W_n^h be the number of comparisons made in the* 1*-stage of* $(h,1)$*-Shell Sort to sort n random keys, so that $W_n^h = n + I_n^h$. Then*

$$\frac{W_n^h}{(n/h)^{3/2}} \xrightarrow{\mathcal{D}} W^h \equiv \sum_{r<s} \int_0^1 |B^r(t) - B^s(t)|\, dt,$$

where the sum extends over all pairs (r,s) *with* $1 \le r < s \le h$.

PROOF. We show first that for the constructed versions with the correspondingly constructed versions $\tilde{F}^1_{n/h}, \ldots, \tilde{F}^h_{n/h}$ of the empirical distributions $F^1_{n/h}, \ldots, F^h_{n/h}$, the random $\binom{h}{2}$-vector

$$\underline{\tilde{Z}}_{n/h} \equiv \left(\int_0^1 |\tilde{D}^{1,2}_{n/h}(t)|\, d\tilde{F}^1_{n/h}(t), \ldots, \int_0^1 |\tilde{D}^{h-1,h}_{n/h}(t)|\, d\tilde{F}^h_{n/h}(t) \right)$$

converges almost surely to $\underline{\tilde{Z}}$, where

$$\underline{\tilde{Z}} \equiv \left(\int_0^1 |\tilde{D}^{1,2}(t)|\, dt, \ldots, \int_0^1 |\tilde{D}^{h-1,h}(t)|\, dt \right).$$

It suffices to prove that each component converges, and, by symmetry, this holds if we show that it holds for the first component. However, this follows exactly as in Section 3, using the decomposition given in (4) and the a.s. continuity and boundedness of the function $|\tilde{D}|$.

It follows that the sum of the components of $\underline{\tilde{Z}}_{n/h}$ converges a.s. to the sum of the components of $\underline{\tilde{Z}}$, and this implies that

$$\frac{I_n^h}{(n/h)^{3/2}} \xrightarrow{\mathcal{D}} W^h,$$

as claimed. □

To the best of our knowledge, the distribution of W^h for $h > 2$ is completely unknown. Of course the summands that comprise W^h have known distributions, but the (i, j)- and (r, s)-summands will be correlated if the pairs (i, j) and (r, s) share an integer. The first moment, which again was found by Knuth (1973), follows easily from the representation. We do not have a closed form for the second moment but it can be evaluated numerically.

COROLLARY 3.

(a) *The average number of comparisons made by the* 1*-stage of* $(h,1)$*-Shell Sort is asyptotically equivalent to*

$$\left(\frac{n}{h}\right)^{3/2} \binom{h}{2} \frac{\sqrt{\pi}}{4}.$$

(b) *The variance of the number of comparisons is asymptotically equivalent to*

$$\left(\frac{n}{h}\right)^{3} \left\{ \binom{h}{2} \left(\frac{7}{30} - \frac{\pi}{16} \right) + h(h-1)(h-2) \left(C - \frac{\pi}{16} \right) \right\},$$

where

$$C \equiv \mathbf{E} \left\{ \int_0^1 |B_1(t) - B_2(t)|\, dt \int_0^1 |B_1(t) - B_3(t)|\, dt \right\}$$

and $B_1(t), B_2(t), B_3(t)$ *are independent (standard) Brownian bridges. The value of* C *is numerically determined to equal* 0.2051.

PROOF. (a) follows immediately from the fact that W^h is the sum of $\binom{h}{2}$ identically distributed terms. From the arguments leading to Corollary 1, the mean of each term is $\sqrt{\pi}/4$.

For (b), the variance of W^h is the sum of $\binom{h}{2}$ terms equal to

$$\mathrm{Var} \left\{ \int_0^1 |B_1(t) - B_2(t)|\, dt \right\},$$

which from Shepp's (1982) results is $7/30 - \pi/16$, plus the covariance terms. For pairs (i, j) and (r, s), the covariance will be zero unless there is a common integer in (i, j) and (r, s). It is easy to check that there are $h(h-1)(h-2)$ such terms, and the common covariance is given by $C - \pi/16$. □

We note a curious consequence of Corollary 3. Although as h increases, the mean of the number of comparisons grows as $\sqrt{h}$, the variance grows slowly with h to a limiting value of about $0.0088n^3$.

The evaluation of C merits some comment. As noted previously, the Brownian bridge is a Gaussian process, and it is well known that for $s < t$, $\mathrm{Cov}(B(s), B(t)) = s(1-t)$. Thus C can be evaluated as a double integral once it is known how to calculate $E(|X||Y|)$, where (X, Y) is bivariate normal with mean (0, 0), unit variances, and correlation ρ. Several approaches are available for this calculation (see Wellner and Smythe, 2001). It can be shown that

$$E(|X||Y|) = \frac{2}{\pi}\{\rho \arcsin(\rho) + \sqrt{1-\rho^2}\},$$

from which evaluation of the covariance can be reduced to a single integration. Alternatively, we can represent $E(|X||Y|)$ as

$$E(|X||Y|) = \frac{2}{\sqrt{\pi}}(1-\rho^2)^{3/2} \sum_{k=0}^{\infty} \rho^{2k} \frac{\Gamma(k+1)}{\Gamma(k+\frac{1}{2})}.$$

The integrated terms of the series decrease rapidly, permitting accurate numerical calculation of C.

For the sake of completeness, we record the analog of Corollary 2 for $(h, 1)$-Shell Sort:

COROLLARY 4.

(a) *Let S_n be the number of comparisons made by* $(h,1)$-*Shell Sort to sort n random keys. Then*

$$\frac{S_n - n^2/4h}{n^{3/2}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{36h^2}\right) \oplus \frac{W^h}{h^{3/2}},$$

where W^h is given in Theorem 2.

(b)

$$E(S_n) = \frac{n^2}{4h} + \frac{\sqrt{\pi}}{4}\binom{h}{2}\left(\frac{n}{h}\right)^{3/2} + o(n^{3/2}).$$

5. A Variation of $(h, 1)$-Shell Sort. We describe briefly here a variation of $(h, 1)$-Shell Sort that seems to be of at least theoretical interest. For clarity, we first describe the method for $h = 3$; the extension to $h > 3$ is fairly obvious.

We begin by performing the 3-sort in the usual way. This results in three sorted subarrays, which we denote $X_{(1)}, \ldots, X_{(n/3)}$, $Y_{(1)}, \ldots, Y_{(n/3)}$, and $Z_{(1)}, \ldots, Z_{(n/3)}$. Now order the Y's and Z's by performing an ordinary insertion sort with these two subarrays. We then have one sorted list of $n/3$ numbers (the X's) and one sorted list of $2n/3$ numbers (which we call the Y–Z list).

Next we do an insertion sort with the X list and the Y–Z list, as follows: start with $X_{(1)}$, then insert consecutively the first two numbers in the ordered Y–Z list; next insert $X_{(2)}$ into this sorted list of three numbers, then insert the next two numbers in the Y–Z list, and so forth until the sort is complete.

Let $I_{n,1}$ and $I_{n,2}$ denote, respectively, the number of inversions encountered in ordering the Y–Z list and the number of inversions in sorting the X's into the Y–Z list. Then $I_{n,1}$ and $I_{n,2}$ are independent, because the former depends only on the relative position of the Y's and Z's, whereas the latter depends only on the position of the $X_{(i)}$ with respect to the sorted Y–Z list. From Lemma 1, it follows that

$$I_{n,1} = \sum_{j=1}^{n/3} T_{n,1}^{(j)},$$

where

$$T_{n,1}^{(j)} = \left| \sum_{i=1}^{n/3} [\mathbf{1}_{\{Z_i < Y_j\}} + \mathbf{1}_{\{Y_i < Y_j\}}] \right|.$$

By reasoning almost identical to that of Lemma 1, we further derive that

$$I_{n,2} = \sum_{j=1}^{n/3} T_{n,2}^{(j)},$$

where

$$T_{n,2}^{(j)} = \left| \sum_{i=1}^{n/3} \mathbf{1}_{\{Y_i < X_j\}} + \mathbf{1}_{\{Z_i < X_j\}} - 2\mathbf{1}_{\{X_i < X_j\}} \right|.$$

From the triangle inequality, we get

$$I_{n,1} + I_{n,2} \le I_n^3,$$

where I_n^3 is the number of comparisons given in Section 4 for (3, 1)-Shell Sort. Proceeding as in Section 3, we find that

$$\frac{I_{n,2}}{(n/3)^{3/2}} = \int_0^1 \sqrt{n/3}\,|F_{n/3}^z(t) + F_{n/3}^y(t) - 2F_{n/3}^x(t)|\, dF_{n/3}^x(t) + o_p(1),$$

where $F_{n/3}^x$, $F_{n/3}^y$, $F_{n/3}^z$ are, respectively, the empirical distribution functions of the X's, Y's, and Z's. An argument very similar to that of Theorem 2 shows that

$$\frac{I_{n,2}}{(n/3)^{3/2}} \xrightarrow{\mathcal{D}} \int_0^1 |B_3(t) + B_2(t) - 2B_1(t)|\, dt, \tag{5}$$

where $B_1(t)$, $B_2(t)$, $B_3(t)$ are three mutually independent Brownian bridges. By the properties of the Brownian bridge, the integrand in (5) has the distribution of $\sqrt{6}|B(t)|$, where $B(t)$ is a standard Brownian bridge, so that

$$\frac{I_{n,2}}{(n/3)^{3/2}} \xrightarrow{\mathcal{D}} \sqrt{6}\int_0^1 |B(t)|\,dt.$$

It also follows from the results of Section 3 that

$$\frac{I_{n,1}}{(n/3)^{3/2}} \xrightarrow{\mathcal{D}} \sqrt{2}\int_0^1 |B^*(t)|\,dt,$$

where $B^*(t)$ is a Brownian bridge independent of $B(t)$, so the total number of comparisons in this case converges to a convolution of the distribution of the limit given in Theorem 1. If $W^*_{n,3}$ denotes the number of comparisons, then

$$\frac{W^*_{n,3}}{(n/3)^{3/2}} \xrightarrow{\mathcal{D}} \sqrt{2}Z \oplus \sqrt{6}Z^*,$$

where Z and Z^* are independent copies of $\int_0^1 |B(t)|\,dt$. The asymptotic mean of the number of comparisons for this modified (3, 1)-Shell Sort is then given by

$$\frac{n^{3/2}\sqrt{\pi}}{4}\left(\frac{1}{3^{3/2}} + \frac{1}{3}\right),$$

which is about 0.905 times the mean given in Corollary 2 for “real” (3, 1)-Shell Sort. The asymptotic variance for the modified method is

$$\left(\frac{n}{3}\right)^{3/2}\left(\frac{7}{30} - \frac{\pi}{16}\right)[1+3],$$

about 0.79 times the variance for (3, 1)-Shell Sort.

The extension of this variation to $h > 3$ is straightforward. With notation as in Section 4, the h-sort is performed as previously, resulting in h ordered lists $\{X^1_j\}, \ldots, \{X^h_j\}$. At the next step, the variation we consider first performs an insertion sort to order the X^{h-1} and X^h lists. Then the X^{h-2} list is sorted together with the combined X^{h-1} and X^h lists, just as the X list was sorted with the Y–Z list in the case of $h = 3$. Next the X^{h-3} list is sorted with the combined X^{h-2}–X^{h-1}–X^h list, by inserting first $X^{h-3}_{(1)}$, then the smallest three numbers from the combined list, then $X^{h-3}_{(2)}$, and so forth. When this sort is complete, the X^{h-4} list is sorted with this combined list, and so on until the list is completely sorted. In this process the number of comparisons required at each stage is independent of the other stages, so the asymptotic distribution of the total number of comparisons is a convolution of the distributions of the numbers at each stage. At the stage where the X^k list is inserted, the number of comparisons $I_{n,h-k}$ satisfies, in the notation of Section 4,

$$\frac{I_{n,h-k}}{(n/h)^{3/2}} = \int_0^1 \sqrt{n/h}\,|F^h_{n/h}(t) + \cdots + F^{k+1}_{n/h}(t) - (h-1)F^k_{n/h}|\,dF^k_{n/h}(t) + o_p(1),$$

and

$$\frac{I_{n,h-k}}{(n/h)^{3/2}} \xrightarrow{\mathcal{D}} \int_0^1 |B_h(t) + \cdots + B_{k+1}(t) - (h-1)B_k(t)|\,dt \tag{6}$$

by an argument similar to that of Section 4, where $B_k(t), \ldots, B_h(t)$ are independent Brownian bridges. The integrand of (6) has the distribution of $\sqrt{(h-k)+(h-1)^2}|B(t)|$ where $B(t)$ is a standard Brownian motion, and we are led to the following result:

THEOREM 3. *Let $W^*_{n,h}$ denote the number of comparisons made in the* 1*-stage of the variation of* $(h, 1)$*-Shell Sort to sort n random keys. Then*

$$\frac{W^*_{n,h}}{(n/h)^{3/2}} \xrightarrow{\mathcal{D}} \sqrt{2}Z^{(1)} \oplus \sqrt{6}Z^{(2)} \oplus \cdots \oplus \sqrt{h(h-1)}Z^{(h-1)},$$

where $Z^{(1)}, \ldots, Z^{(h-1)}$ are independent copies of $\int_0^1 |B(t)|\, dt$.

From Theorem 3 it is easy to determine the asymptotic moments for the variation of Shell Sort.

COROLLARY 5.

(a) *The average number of comparisons made by the* 1*-stage of the variation of Shell Sort is asymptotically equivalent to*

$$\left(\frac{n}{h}\right)^{3/2} \frac{\sqrt{\pi}}{4\sqrt{2}} \sum_{j=1}^{h} \sqrt{j(j-1)}.$$

(b) *The variance of the number of comparisons for the variation is asymptotically equivalent to*

$$\frac{n^3}{6}\left(\frac{7}{30} - \frac{\pi}{16}\right)\left(1 - \frac{1}{h^2}\right).$$

Comparing the results of Corollary 3 with those of Corollary 2 for "real" $(h, 1)$-Shell Sort, the asymptotic mean of the variation grows at rate

$$n^{3/2}\frac{\sqrt{\pi}}{4}\frac{\sqrt{h}}{2\sqrt{2}} + O(h^{-1/2}),$$

less than that of the "real" version by a factor of $1/\sqrt{2}$. The variance exhibits the same curious feature of growing slowly with h, but to a limiting value of $0.00616n^3$, compared with $0.0088n^3$ for the "real" version.

6. A Remark on $(\boldsymbol{h}, \boldsymbol{k}, \mathbf{1})$-Shell Sort. In the case where k divides h, the method of Section 4 can be used to find the limiting distribution of $(h, k, 1)$-Shell Sort. The h-stage performs an h-sort, as before, resulting in h ordered lists of length n/h. The k-stage then produces k sorted lists, each formed by an insertion sort on h/k of the h ordered lists. The asymptotic number of comparisons involved in this step is given in Theorem 2 (or even Theorem 1 if $h/k = 2$). The asymptotic number of comparisons in the 1-stage is again found from Theorem 2, and the asymptotic distribution of the total number of comparisons in the k-stage and 1-stage is the convolution of the distributions of the two stages.

For example, in the dyadic form of Shell Sort, with (in the notation of Section 2) $t_k = 2^k$, the limiting distribution of the number of comparisons can be characterized. In (4, 2, 1)-Shell Sort the average number of comparisons in the merging stages (not counting the initial sorts) is $(\sqrt{\pi}/16)n^{3/2}(1+\sqrt{2})$, while in (4, 1)-Shell Sort the average is $(\sqrt{\pi}/16)n^{3/2}(3)$, giving an "efficiency factor" of $(1+\sqrt{2})/3$. In general, using the sequence $(2^k, 2^{k-1}, \ldots, 1)$ compared with the $(2^k, 1)$-Shell Sort gives an "efficiency factor" of

$$\frac{2^{(k/2)-1}(1+\sqrt{2})}{2^k - 1}.$$

Because of poor worst-case behavior, the case where k divides h is generally not of practical interest, however.

7. Some Monte-Carlo Results. Here we present the results of some Monte-Carlo experiments to confirm the limit theorems presented in the previous sections. Figures 1–3 present the empirical distributions of 5000 replications of the standardized number of comparisons S_n for $n = 500, 5000$, and ∞. Instead of standardizing as in Corollaries 2 and 4, we have chosen to present the empirical distributions of

$$\frac{S_n - n^2/(4h)}{(n/h)^{3/2}} \to_d N\left(0, \frac{h}{36}\right) + W^h.$$

The Monte-Carlo experiments were carried out for $n = 500, 5000$ using a C-program by David Butler at Oregon State University, and the Monte-Carlo experiments for the limiting Brownian bridge process limits were done using Mathematica by Wellner at University of Washington. The latter involved approximating Brownian bridge processes by (tied-down) partial sum processes of pseudo-random Gaussian variables using a grid of

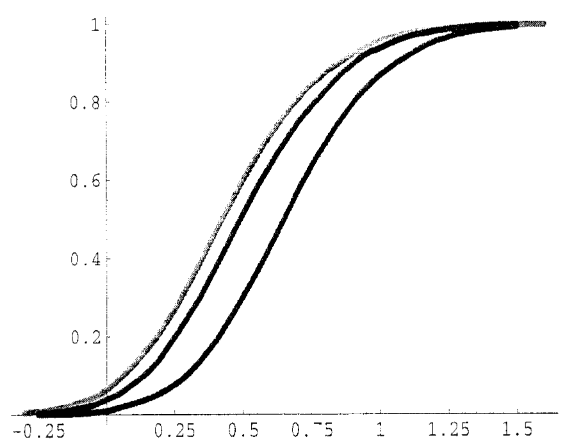

Fig. 1. Empirical distributions for (2,1)-Shellsort, $n = \infty, 5000, 500$.

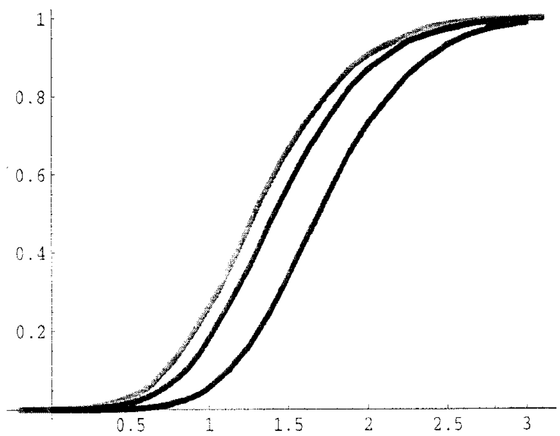

Fig. 2. Empirical distributions for (3,1)-Shellsort, $n = \infty, 5000, 500$.

10^4 points on the interval [0, 1]. (Empirical distributions of this Monte-Carlo method were checked for agreement with the theoretically known distribution of $\int_0^1 |B(t)|\, dt$ given in Johnson and Killen (1983).) The computer codes for both of these programs are available from the authors.

In Figures 1–3 the plot furthest to the right is for $n = 500$, the plot for $n = 5000$ is in the middle, and the plot of the asymptotic distribution (for "$n = \infty$") is the furthest to the left. This is in agreement with the empirical and theoretical means shown in Table 1. Table 2 gives a comparison of empirical and theoretical variances, indicating that variances decrease slightly with n for all three situations. The shapes of the empirical distributions for $n = 500, 5000$, and ∞ seem to be very similar in all three situations, and the limit distribution is being converged to from below (i.e., from the right). This is probably due to terms of order $o(n^{3/2})$ which are ignored in the analysis.

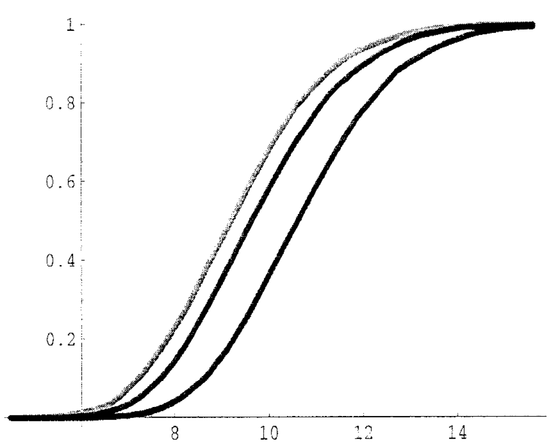

Fig. 3. Empirical distributions for (7,1)-Shellsort, $n = \infty, 5000, 500$.

Table 1. Theoretical and empirical (Monte-Carlo) means; 5000 replications.

n	(2, 1)	(3, 1)	(7, 1)
Theory ∞	0.443	1.329	9.305
Empirical ∞	0.444	1.328	9.297
Empirical 5000	0.509	1.450	9.765
Empirical 500	0.662	1.729	10.717

Table 2. Theoretical and empirical (Monte-Carlo) variances; 5000 replications.

n	(2, 1)	(3, 1)	(7, 1)
Theory ∞	0.0925	0.2468	2.8087
Empirical ∞	0.0971	0.2466	2.7875
Empirical 5000	0.0929	0.2480	2.8643
Empirical 500	0.0925	0.2522	2.9119

Acknowledgments. The authors thank Hosam Mahmoud for demystifying Shell Sort and providing editorial advice on Sections 1 and 2; and Michael Perlman and Svante Janson for help with the formulas for $E|X||Y|$ in Section 4.

References

Dudley, R. M. (1997). Empirical processes and p-variation. *Festschrift for Lucien Le Cam*, D. Pollard, E. Torgersen, and G. Yang, eds. Springer-Verlag, New York, pp. 219–233.

Johnson, B. McK., and Killeen, T. (1983). An explicit formula for the C.D.F. of the L_1 norm of the Brownian bridge. *Annals of Probability* **11**, 807–808.

Knuth, D. (1973). *The Art of Computer Programming*, Volume 3: Sorting and Searching. Addison-Wesley, Reading, Massachusetts.

Lent, J., and Mahmoud, H. (1996). On tree-growing search strategies. *The Annals of Applied Probability* **6**, 1284–1302.

Louchard, G. (1986). Brownian motion and algorithm complexity. *BIT* **26**, 17–34.

Mahmoud, H. M. (2000). *Sorting*: *A Distribution Theory*. Wiley, New York.

Sedgewick, R. (1988). *Algorithms*, 2nd edn. Addison-Wesley, Reading, Massachusetts.

Sedgewick, R. (1996). Analysis of Shellsort and related algorithms. *Algorithms - ESA* '96: *Fourth Annual European Symposium on Algorithms*, *Barcelona*, *Spain*, *Sept*. 25–27, 1996, J. Diaz and M. Serna, eds. Springer-Verlag, Berlin, New York, pp. 1–11.

Sedgewick, R., and Flajolet, P. (1996). *An Introduction to the Analysis of Algorithms*. Addison-Wesley, Reading, Massachusetts.

Shepp, L. (1982). On the integral of the absolute value of the pinned Wiener process. *Annals of Probability* **10**, 234–239.

Shorack, G. R., and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

Wellner, J. A., and Smythe, R. T. (2001). Computing the covariance of two Brownian area integrals. *Statistica Neerlandica*, Vol. 55, to appear.